

# Integração de sistemas para predição de deslizamentos de terra baseada em aprendizado de máquina

Laedson Silva Pedreira<sup>1</sup>, Maria do S. C. São Mateus<sup>1</sup>, Rodrigo Tripodi Calumby<sup>1</sup>

<sup>1</sup>Universidade Estadual de Feira de Santana (UEFS)

Av. Transnordestina, s/n - Feira de Santana, Novo Horizonte - BA, 44036-900

laedsonsilva@hotmail.com, socorro@uefs.br, rtcalumby@uefs.br

**Abstract.** *Landslides are one of the main causes of natural disasters, resulting in material damage and fatalities. Therefore, an intelligent system capable of predicting these events would be useful to support public authorities in the decision-making process and situational management in smart cities. This work performed the integration of multiple systems required for facing the problem and applied data mining and machine learning techniques to build an integrated database and predict rainfall-induced landslides. The experiments showed promising results, with roughly 90% of the landslides records being correctly predicted.*

**Resumo.** *Os deslizamentos de terra constituem um dos principais fenômenos causadores de desastres naturais, provocando prejuízos materiais e vítimas fatais. Por isso, um sistema inteligente capaz de prever esses eventos seria útil para suporte ao poder público no processo de tomada de decisão e no gerenciamento situacional em cidades inteligentes. Este trabalho realizou a integração de múltiplos sistemas necessários para enfrentar o problema e utilizou técnicas de mineração de dados e aprendizado de máquina na construção de uma base de dados integrada e na predição de deslizamentos de terra induzidos por chuvas. Os experimentos demonstraram resultados promissores, sendo corretamente preditos cerca de 90% dos registros de deslizamentos.*

## 1. Introdução

Os deslizamentos de terra ocorrem por todo planeta e vêm provocando, todos os anos, inúmeros problemas para a sociedade, tais como a destruição de edificações e equipamentos urbanos, prejuízos em atividades produtivas, impactos ambientais e fazendo um grande número de vítimas fatais [IBGE 2019]. Eles podem ser desencadeados por inúmeros fatores naturais, como precipitação pluviométrica e terremotos, ou por razões antropogênicas, como corte e sobrecarga de encostas. Dentre esses fatores, os deslizamentos de terra são potencializados, principalmente, pela ação das águas das chuvas, que infiltram no solo das encostas, provocando a redução da sua estabilidade e, consequentemente, induzindo os deslizamentos [Farah 2003]. Impedir que esses eventos ocorram foge da capacidade humana, sendo necessária a adoção de medidas de prevenção que minimizem os impactos causados por eles [Kobiyama et al. 2006].

Nesse contexto, a previsão de deslizamentos de terra pode ser extremamente útil para minimizar o número de vítimas e reduzir os danos materiais, de forma a dar suporte ao poder público no processo de tomada de decisão e no gerenciamento situacional, a

fim de reduzir os danos causados por esses fenômenos. Contudo, o desenvolvimento de uma abordagem para previsão de deslizamentos de terra é muito desafiador. Primeiramente devido à baixa disponibilidade de dados integrados e em larga escala, além de que uma estimativa precisa envolve um número considerável de variáveis inter-relacionadas, o que torna a previsão de deslizamentos de terra um problema não trivial com amostras e separação não linear no espaço.

Métodos de engenharia geotécnica são utilizados para avaliar a relação entre a resistência ao cisalhamento do solo e as tensões cisalhantes mobilizadoras que determinam a estabilidade, ou não, de uma encosta. No entanto, esses métodos têm uma grande limitação, pois dependem da qualidade da investigação em um local específico, o que pode ser muito caro. Ademais, a escassez de medições consideravelmente detalhadas e em tempo real das condições do solo, do maciço rochoso e das águas subterrâneas, impede a previsão precisa dos deslizamentos de terra.

Desta forma, os pesquisadores estão explorando cada vez mais técnicas relacionadas à utilização de Mineração de Dados (MD) e Aprendizado de Máquina (AM) na tentativa de aproximar a ocorrência de deslizamentos futuros a padrões de distribuição anteriores, em função do grande potencial dessas técnicas em descobrir relações complexas entre os dados com múltiplos componentes associados [Korup and Stolle 2014]. Entretanto, são escassos os estudos que buscam prever deslizamentos de modo individualizado e com relação temporal, devido à baixa disponibilidade de dados precisos de deslizamentos (geolocalização e data), fatores desencadeantes (precipitação pluviométricas) e fatores de controle (geomorfologia, geologia etc). Com isso, é notório que essa área de aplicação encontra-se em desenvolvimento, com vários desafios e limitações a serem superadas.

Nesse sentido, em função da complexidade do contexto apresentado e dos desafios computacionais a serem superados, a construção de um modelo preditivo de deslizamentos de terra passa pela obtenção de dados, garantia e melhoria de qualidade dos dados, integração dos dados, escolha de algoritmos, construção, otimização e validação de modelos eficazes. Este artigo propõe a construção de uma base de dados a partir de dados geográficos e temporais e apresenta os resultados de algoritmos de classificação para identificar eventos de deslizamentos de terra induzidos por chuvas.

## 2. Trabalhos Relacionados

Em [Souza and Ebecken 2012], foi desenvolvida uma metodologia para prever deslizamentos de terra induzidos por chuvas na cidade do Rio de Janeiro. O inventário de deslizamentos continha 1.033 amostras que foram categorizados por bairro. Foram geradas 233 amostras de não escorregamento. Foram desenvolvidos três modelos, com três abordagens: (1) previsão de deslizamentos; (2) extração de regras de associação; e (3) previsão de chuvas. A primeira abordagem utilizou RNA e Regras de Classificação para extrair um conjunto de regras interessantes com base em dois índices: suporte e confiança. A segunda abordagem utilizou o algoritmo *Apriori* para extrair regras de associação da base de dados. Por fim, a terceira abordagem gerou um modelo para prever precipitação pluviométrica. Com isso, tentou-se realizar a previsão de deslizamentos associando às regras com a precipitação prevista. O percentual de amostras usadas para treino, teste e a validação foi de 80, 10 e 10%, respectivamente. A RNA obteve uma taxa de acerto de 94,1% para não ocorrência, 80,7% para um estado de pânico e 72,4% para deslizamentos.

Em [Farahmand and Aghakouchak 2013], foi desenvolvido um modelo de previsão de deslizamentos de terra usando dados de precipitação por satélite, mapas de uso do solo e informações topográficas e, para isso, utilizaram o algoritmo *Support Vector Machines*. Neste modelo foram utilizados dados de precipitação pluviométricas, dados de solo e um inventário contendo 581 eventos de ocorrências. Foram geradas 5.810 amostras de não deslizamento de forma aleatória. Para o treinamento e validação, foi realizada uma subamostragem aleatória com 100 repetições, sendo os dados divididos em 70% para treino e 30% para validação. O erro médio de 100 iterações de previsão de deslizamento foi estimado em aproximadamente 7%, sendo 2% de deslizamentos falsos, e aproximadamente 7% de eventos de deslizamentos classificados como não deslizamentos.

Em [Tehrani et al. 2019], foram implementados os algoritmos Regressão Logística para classificar eventos de deslizamentos e não deslizamentos. Os autores alcançaram uma AUC superior a 0,88, e utilizaram uma base de dados global com 4.542 deslizamentos de terra. Contudo, relataram que a localização dos eventos apresentava grandes incertezas. As amostras de não ocorrências foram geradas aplicando uma fração aleatória (menos de 0,5) às características topográficas e pluviométricas de casos de deslizamentos. Foi considerado o acumulado de até 11 dias anterior ao deslizamento. Ao todo foram criadas 4.542 amostras de não deslizamento. Com a técnica de *Hold-out*, dividiram a base de dados em treino (67%) e teste (33%).

[Farahmand and Aghakouchak 2013] e [Souza and Ebecken 2012] esbarraram na disponibilidade e qualidade dos dados. Utilizaram apenas 581 e 1.033 amostras de deslizamento de terra, respectivamente e, mencionam a baixa qualidade da geolocalização dos ocorrências. No caso de [Souza and Ebecken 2012], tinham disponível apenas o bairro onde ocorreu o evento, enquanto [Farahmand and Aghakouchak 2013] citam imprecisões de 1 a 5 km na localização dos eventos. Os autores não conseguiram contornar o problema do desbalanceamento de classes. Apesar desse desequilíbrio, não utilizaram nenhuma estratégia de amostragem, o que pode gerar uma configuração tendenciosa, devido à grande desproporcionalidade entre as classes, tornando fácil para o modelo fornecer a saída correta, e com isso maximizando as medidas de desempenho.

Ainda, esses trabalhos não mencionaram a utilização de técnicas de validação como o *Cross-Validation*, que permite estimar com maior rigor a capacidade de generalização dos modelos. Já o trabalho de [Tehrani et al. 2019] propôs um modelo de previsão de deslizamentos de terra em escala global, sendo uma abordagem de difícil aplicação prática. Ademais, o método de geração de amostras negativas pode impor um viés aos conjuntos de treinamento e teste.

### 3. Métodos

#### 3.1. Coleta dos dados

Os conjuntos de dados coletados a partir de sistemas de informação de diversas instituições foram:

1. O **Inventário de deslizamentos** foi obtido dos registros da Defesa Civil de Salvador (CODESAL), correspondendo 4.892 eventos no período de janeiro de 2004 a junho de 2021. Cada registro é constituído por: número do processo, código da vistoria, data e hora da abertura do processo, data e hora da vistoria, tipo de ocorrência, bairro, prefeitura-bairro, causa, descrição e coordenadas UTM.

2. Os **dados Geotécnicos e Geológicos** foram obtidos dos registros do Laboratório de Geotecnia da Universidade Federal da Bahia (UFBA) e da Companhia de Pesquisa de Recursos Minerais (CPRM), incluindo as propriedades do solo do município de Salvador (coesão, ângulo de atrito e peso específico), domínio geológico, relevo e composição do solo.
3. Os **dados de Precipitações Pluviométricas** complementares foram coletados de sistemas de quatro instituições. Dados de 31 estações pluviométricas foram coletados por meio da CODESAL, abrangendo o período entre outubro de 2016 a outubro de 2021. Dados de 20 estações pluviométricas foram coletados do sistema disponibilizado pelo Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN)<sup>1</sup>, correspondendo ao período desde setembro de 2013. O sistema do Instituto Nacional de Meteorologia (INMET) foi utilizado para coleta de dados de 3 estações pluviométricas convencionais com registros diários de precipitações. Estão disponíveis registros obtidos entre agosto de 1963 e outubro de 2021. Dados de outras 5 estações foram obtidos do sistema do Instituto do Meio Ambiente e Recursos Hídricos (INEMA), compreendendo o período entre 1998 e 2016. Todos estes sistemas foram integrados devido à grande influência da precipitação no processo de deslizamento de terra, buscando-se maximizar a quantidade e cobertura espacial dos dados de precipitação.
4. Os **dados Geomorfométricos** foram obtidos do banco de dados do sistema Topodata [Brasil 2008], que oferece informações de variáveis geomorfométricas locais em cobertura nacional. Deste sistema, foram incorporados o Modelo Digital de Elevação (MDE) e sua derivação de declividade.

### 3.2. Integração de dados

Como resultado da integração, objetivou-se a criação de uma base de dados preparada para aplicação de técnicas de MD e AM. Grande parte dos dados adquiridos são do tipo geoespaciais, processados com apoio do software QGIS versão 3.22.3<sup>2</sup>. Desta forma, os dados de deslizamentos foram integrados com os dados geomorfométricos, geológicos e áreas de riscos, considerando-se a localização geoespacial.

Como os dados incluem a latitude e a longitude para cada estação pluviométrica, foi possível determinar a distância de cada escorregamento para cada estação. Deste modo, a precipitação diária foi obtida por meio da estação mais próxima que possui dados disponíveis para a data do deslizamento e para período anterior.

### 3.3. Pré-processamento de Dados

O primeiro processo realizado foi a remoção de registros duplicados. Em seguida foram eliminados os registros de deslizamentos que não foram induzidos por chuvas ou que foram categorizados erroneamente como deslizamentos. Por fim, foram excluídos os registros que possuíam valores ausentes nos atributos de coesão, ângulo de atrito e peso específico. Esses valores ausentes têm origem nos deslizamentos que ocorreram em áreas para quais não se tinha as propriedades disponíveis no sistema provedor. Este processo resultou em 1.864 registros descartados, restando 3.028 registros de escorregamentos aptos e consistentes.

<sup>1</sup><http://www2.cemaden.gov.br/mapainterativo/>

<sup>2</sup>[https://www.qgis.org/pt\\_BR/site/forusers/download.html](https://www.qgis.org/pt_BR/site/forusers/download.html)

No processo de engenharia de atributos, para cada deslizamento, foram selecionados os índices de chuvas diárias e criados seus respectivos acumulados. Considerando a data da ocorrência do evento de deslizamento, foram criados 17 índices pluviométricos (dados obtidos do pluviômetro mais próximo), incluindo a precipitação do dia do evento, a precipitação diária de 1 a 8 dias antes do deslizamento e os respectivos acumulados.

Os dados referentes ao mês de ocorrência, domínio geológico, relevo, composição do solo e zonas de riscos foram transformados por meio da técnica de *One-Hot Encoding*, por meio da qual uma variável categórica é convertida em múltiplas variáveis binárias correspondentes aos possíveis valores das variáveis originais.

Para a validação experimental dos modelo preditivos, registros correspondentes à não ocorrência de deslizamentos foram criados e adicionados ao conjunto de amostras. Para isso, gerou-se para cada registro de deslizamento, um outro registro correspondente à uma não ocorrência de deslizamento. Cada amostra de não escorregamento foi coletada um dia antes da ocorrência do evento de deslizamento. Com isso, tem-se as duas classes possíveis: positiva (deslizamento) e negativa (não deslizamento).

### 3.4. Algoritmos de Aprendizado de Máquina (AM)

Para este trabalho foram escolhidos os algoritmos de Árvore de Decisão (AD), *Random Forest* (RF), *Light Gradient Boosting Machine* (LGBM) e Redes Neurais Artificiais (RNA) do tipo *Multilayer Perceptron*. Foram utilizados os métodos implementados em linguagem *Python* por meio da biblioteca *Scikit-learn* e então realizada a avaliação experimental na predição de deslizamentos de terra.

Estes algoritmos supervisionados baseiam-se em abordagens distintas, são amplamente utilizados e bem documentados. Foram escolhidos dois algoritmos com abordagem convencional e dois algoritmos do tipo *ensemble*. Por fim, os algoritmos AD, RF e LGBM ainda não foram explorados para o problema de predição de deslizamentos no tempo e no espaço, conforme o levantamento de trabalhos anteriores, sendo que apresentam resultados promissores para diversos tipos de problemas. Os algoritmos utilizados são brevemente descritos a seguir.

As AD constituem métodos que utilizam-se de uma representação com base em árvores. Para a construção da árvore é feita uma sequência de divisões no conjunto de dados baseadas na seleção de melhores pontos de separação de valores dos atributos e, assim, determina-se a ramificação da árvore até que um determinado critério de parada seja satisfeito. Existe o nó raiz, em que inicia a árvore; os nós de decisão, que estabelecem o caminho da árvore; e os nós terminais/folha, que representam uma classe ou valor contínuo [Breiman et al. 1984].

A RF é um algoritmo do tipo *ensemble* utilizado para classificação e regressão que combina o resultado de um conjunto de árvores de decisão para realizar a predição. Para isso, utiliza métodos como *Bagging* e *Feature Randomness* para construir suas árvores, e dessa forma, cria uma floresta com baixa correlação. A RF utiliza a ideia de que a decisão por voto para previsão final é melhor do que a previsão individual de cada árvore [Breiman 2001].

A rede *Multilayer Perceptron* consiste em uma RNA que possui múltiplas camadas de neurônios que são ligadas entre si por sinapses contendo pesos. Consiste em uma

camada de entrada e uma saída com uma ou mais camadas ocultas. O treinamento desse tipo de RNA é realizado, geralmente, pelo método de *back-propagation*. Essas redes podem lidar com dados que não são linearmente separáveis, ou separáveis por um hiperplano [Haykin 2009].

O LGBM é um *Gradient Boosting Framework* baseado em árvores de decisão. Diferente dos outros modelos baseados em árvores, no LGBM, as árvores crescem na vertical, ou seja, utiliza a estratégia de crescimento de árvore em folha, enquanto os outros algoritmos de *boosting* crescem a árvore horizontalmente, isto é, crescem a profundidade da árvore em nível. O LGBM seleciona a folha de árvore com o melhor ajuste para crescer e assim pode diminuir mais perdas do que um algoritmo com base no crescimento horizontal [Ke et al. 2017].

### 3.5. Protocolo Experimental e Medidas de Eficácia

Preliminarmente, 20% do conjunto de dados foi reservado para testes independentes adicionais e, para isso realizou-se uma amostragem aleatória estratificada. Por sua vez, experimentos usando o protocolo de validação cruzada com 10 *folds* foram realizados sobre os 80% restantes dos dados. A otimização dos hiperparâmetros foi feita através da técnica *grid search*. Com base em uma grade de parâmetros, o *grid search* cria  $n$  combinações, com o objetivo de encontrar a que otimize o desempenho de cada classificador. Para cada combinação criada, ou seja, para cada modelo construído é executado o método de validação cruzada com 10 *folds*, a fim de avaliar o grau e a capacidade de generalização de cada modelo. Para cada iteração da validação cruzada são computadas as medidas de avaliação. Ao término do processo, as performances estimadas de cada modelo, são utilizadas para calcular a performance média. Foi utilizada a medida F1-score (F1) para a escolha do melhor modelo, contudo também foram computados os valores das seguintes medidas de avaliação: acurácia (ACC), *precision*, *recall* e área sob a curva ROC (AUC).

Ao final da execução do *grid search* foi escolhido o modelo que apresentou o melhor desempenho médio. Por fim, esse modelo foi aplicado sobre o conjunto dados reservado para testes adicionais, com o objetivo de verificar a capacidade preditiva desses modelos em dados inéditos adicionais, não utilizados em qualquer etapa anterior de otimização dos modelos.

Este processo foi repetido dez vezes, sendo em cada execução atribuída uma nova semente de aleatoriedade, com o objetivo de embaralhar os dados de maneira distinta da iteração anterior. Desta forma, os conjuntos de dados de treino e teste variaram a cada iteração e em cada rodada da validação cruzada.

## 4. Resultados e Discussões

Considerando as melhores configurações de hiperparâmetros encontradas, foram obtidos os melhores modelos de AM. Os dois modelos de maior AUC foram comparados estatisticamente, com o objetivo de verificar se há diferenças significativas entre os resultados, sendo utilizado o teste não paramétrico de *Wilcoxon's signed-rank test* com confiança de 95%. As Tabelas 1 e 2, apresentam, respectivamente, os resultados da validação cruzada e dos resultados dos teste adicionais.

A avaliação de performance dos modelos selecionados evidenciou um desempenho expressivo, com F1 superior a  $0,761 \pm 0,022$  e AUC superior a  $0,762 \pm 0,020$ ,

para todos os modelos. Os algoritmos LGBM e RF apresentaram resultados ainda melhores, obtendo valores médios de F1 superiores a  $0,865 \pm 0,006$  e AUC superiores a  $0,865 \pm 0,006$ . Ainda, esses algoritmos apresentaram menores dispersões em relação a média quando comparados aos demais, além de equilíbrio entre as medidas de *precision* e *recall*.

De modo geral, observou-se que o algoritmo RF obteve, resultados estatisticamente superiores na maioria das medidas, alcançando  $F1 = 0,887 \pm 0,003$   $AUC = 0,875 \pm 0,003$ . O LGBM, em comparação com a RF, apresentou resultados estatisticamente superiores em termos de *recall*, sendo preditos corretamente cerca de 90% dos deslizamentos registrados e com taxa similar de falsos positivos.

Considerando os dados retidos para testes adicionais, completamente inéditos e que não foram utilizados nem mesmo para otimização dos hiperparâmetros dos modelos, os algoritmos conseguiram repetir o bom desempenho alcançando na validação cruzada, demonstrando boa capacidade de generalização das configurações utilizadas. Além disso, pode-se observar que houve uma maior dispersão dos resultados em relação à média, não havendo o comprometimento da significância estatística.

**Tabela 1. Resultados obtidos na validação cruzada**

Algoritmo	F1	ACC	AUC	Recall	Precision
RNA	$0,770 \pm 0,008$	$0,763 \pm 0,013$	$0,763 \pm 0,013$	$0,792 \pm 0,037$	$0,752 \pm 0,029$
AD	$0,843 \pm 0,002$	$0,829 \pm 0,004$	$0,829 \pm 0,004$	$0,913 \pm 0,012$	$0,783 \pm 0,009$
RF	<b><math>0,877 \pm 0,003^*</math></b>	<b><math>0,875 \pm 0,003^*</math></b>	<b><math>0,875 \pm 0,003^*</math></b>	$0,889 \pm 0,004$	<b><math>0,865 \pm 0,004^*</math></b>
LGBM	$0,875 \pm 0,003$	$0,871 \pm 0,003$	$0,871 \pm 0,003$	<b><math>0,899 \pm 0,003^*</math></b>	$0,852 \pm 0,004$

\* Medida significativamente maior (Wilcoxon signed-rank,  $\alpha < 0,05$ )

**Tabela 2. Resultados de testes adicionais**

Algoritmo	F1	ACC	AUC	Recall	Precision
RNA	$0,761 \pm 0,022$	$0,762 \pm 0,020$	$0,762 \pm 0,020$	$0,783 \pm 0,046$	$0,755 \pm 0,036$
AD	$0,823 \pm 0,013$	$0,824 \pm 0,013$	$0,824 \pm 0,013$	$0,915 \pm 0,007$	$0,775 \pm 0,018$
RF	<b><math>0,869 \pm 0,006^*</math></b>	<b><math>0,869 \pm 0,006</math></b>	<b><math>0,869 \pm 0,006</math></b>	$0,883 \pm 0,014$	<b><math>0,859 \pm 0,008^*</math></b>
LGBM	$0,865 \pm 0,006$	$0,865 \pm 0,006$	$0,865 \pm 0,006$	<b><math>0,893 \pm 0,014^*</math></b>	$0,846 \pm 0,005$

\* Medida significativamente maior (Wilcoxon signed-rank,  $\alpha < 0,05$ )

## 5. Considerações Finais e Trabalhos Futuros

Este estudo realizou a integração de múltiplos sistemas e tipos de dados, com significado importante para a predição de deslizamentos de terra induzidos por chuva. Após todo o processo de ciência de dados, como a coleta, integração, análise e processamento de dados, foi possível aplicar quatro algoritmos de AM e realizar sua avaliação e comparação.

Os modelos de classificação testados e avaliados mostraram-se capazes de encontrar padrões na base de dados construída e com isso podem, com relevante desempenho preditivo, realizar a previsão de deslizamentos de terra induzidos por chuvas. Pode-se constatar que a utilização do AM sobre a integração de múltiplos sistemas pode ser uma solução promissora para auxiliar na tomada de decisões pelos órgãos competentes, conferindo melhorias no processo de monitoramento e redução dos danos causados pelos deslizamentos de terra, de forma que podem ser utilizados no monitoramento em tempo

real das áreas de riscos e na identificação das áreas mais críticas a ocorrências desses fenômenos.

Como trabalhos futuros destaca-se a necessidade de avaliação de diferentes métodos de geração de amostras negativas, bem como a combinação desses métodos. Sugere-se também analisar cenários com índices pluviométricos diários e acumulados com diferentes janelas temporais, a fim de verificar o o impacto do número dias de chuva acumulada na predição de deslizamentos. Além disso, espera-se verificar a eficácia com diferentes períodos de antecedência preditiva e o impacto da imprecisão da precipitação na capacidade preditiva dos algoritmos.

## 6. Agradecimentos

Os autores deste estudo agradecem a CODESAL, ao Laboratório de Geotecnia da UFBA e a CPRM pela atenção e pela disponibilização dos dados utilizados no trabalho.

## Referências

- Brasil (2008). *Instituto Nacional de Pesquisas Espaciais (INPE). Topodata: banco de dados geomorfométricos do Brasil. Variáveis geomorfométricas locais.*
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees*. Taylor & Francis.
- Farah, F. (2003). *Habitação e encostas*, volume volume. Instituto de Pesquisas Tecnológicas (IPT), São Paulo.
- Farahmand, A. and Aghakouchak, A. (2013). A satellite-based global landslide model. *Natural Hazards and Earth System Science*, 13(5):1259–1267.
- Haykin, S. S. (2009). *Neural networks and learning machines*. Pearson Education, third edition.
- IBGE (2019). *Suscetibilidade a deslizamentos do Brasil: primeira aproximação*. Instituto Brasileiro de Geografia e Estatística - IBGE, Rio de Janeiro.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 2017-December(Nips):3147–3155.
- Kobiyama, M., Mendonça, M., Moreno, D. A., Marcelino, I. P. d. O., Marcelino, E. V., Brazetti, L. L. P., Goerl, R. F., Moller, M. G. S. F., and Rudolf, F. d. M. R. (2006). Prevenção de desastres naturais - - Conceitos Básicos. page 109.
- Korup, O. and Stolle, A. (2014). Landslide prediction from machine learning. *Geology Today*, 30.
- Souza, F. T. d. and Ebecken, N. F. (2012). A Data Based Model to Predict Landslide Induced by Rainfall in Rio de Janeiro City. *Geotechnical and Geological Engineering*, 30(1):85–94.
- Tehrani, F. S., Santinelli, G., and Herrera, M. (2019). A framework for predicting rainfall-induced landslides using machine learning methods. *17th European Conference on Soil Mechanics and Geotechnical Engineering, ECSMGE 2019 - Proceedings*.